

World's Most Comprehensive Protein Search Engine

Transforming Protein Discovery Processes for Industrial Biocatalysis and Pharma

In the rapidly evolving field of bioinformatics, the ability to access and analyze protein data is crucial for scientific advancement. While established tools have been instrumental in allowing researchers to find protein sequences, there are limitations, particularly when it comes to accessing novel data. Proteineer has developed a new search engine designed to bridge these gaps and provide a more comprehensive approach to protein discovery.

CHEManager: What inspired you to create a new protein search engine?

Johannes Kabisch: The realization that there are huge amounts of untapped sequence data besides the most commonly used Genbank. Our vision with Proteineer became to expand the accessibility for our customers to these already existing, yet very difficult to access data. Our search engine goes beyond just sequence similarity, we also use structural similarity and search through many data sources including short read archives thus including novel proteins that have not been previously discoverable. In short, with our approach we find far more protein sequences than any other search engine.

How many proteins do you typically find in a search?

Joe Heenan: The number of proteins we identify depends on the search criteria, but it's usually over 100,000 results. While this number seems large and unmanageable for testing, it is good to have more results to filter them for preferable properties.

Once we find a protein, we predict its structure and various useful properties, such as thermostability, optimal operating temperatures, and expected expression levels in *E. coli*. We as well assess the structural simi-



Aron Eiermann, CTO, Proteineer

larity to the query protein and perform clustering. This in-depth analysis provides researchers with a comprehensive understanding of the search space and a wide variety in the selection of proteins to be tested in the laboratory.

What challenges did you face in developing and providing your services?

Aron Eiermann: One of the major challenges was scaling our processes, which require really massive computational power, to be cost-effective for a larger customer base. This was solved by collaborating with the largest computational providers like OVH, Google, Amazon and NVIDIA. Another unexpected challenge was that the customers were overwhelmed by the amount of sequence data we could provide them, which was more than is readily testable in a lab. We have tackled this by developing a software called GeneStore which allows our customers to use their expertise about their proteins to select a diverse panel they want to produce and characterize.

How does that work?

J. Kabisch: The GeneStore is our web application with which we present search results, allowing customers to explore protein properties, including structural and docking results. They can filter these results based on various properties or start new searches.



Johannes Kabisch, CSO, Proteineer

It's a comprehensive tool for managing and analyzing protein data.

What are Proteineer's future plans?

A. Eiermann: We've recently built a robust docking pipeline that can also be used by biologists who are not experts in structural biology. We will integrate this into the GeneStore to allow our users to filter proteins based on their binding properties. We believe that this will enable our clients to select even better candidates for laboratory testing.

J. Heenan: We are also working on integrating language models to provide in-silico data and use our extensive training data to improve our offering.

J. Kabisch: Another exciting potential we see is the expansion of our services to support clients in obtaining freedom to operate and to help them file robust and comprehensive patent applications using the sequence space we offer.

What is your business model?

J. Heenan: We operate on a Software as a Service (SaaS) model, offering subscriptions to GeneStore. We also take on specific bioinformatics challenges on a contract basis. With our services, we support biologists by providing them with new proteins and ideas for their research and helping them to better understand their proteins.



Joseph Heenan, CEO, Proteineer

PERSONAL PROFILE

Aron Eiermann (CTO) has worked in bioinformatics (robotics and classical artificial intelligence) at TU Darmstadt for three years. He has supervised two student software development teams at TU Darmstadt, is a board member and co-founder of Freie Netze Südhessen (association for the promotion of free access to internet). Aron has experience in system integration and scaling as well as developing search algorithms for huge data sets (bigdata).

Johannes Kabisch (CSO) teaches and researches at NTNU, Norway's largest university, with so-called microbial cell factories to establish a bio-economy. Johannes brings his knowledge of DNA sequencing and protein production to Proteineer. During his research activities in Germany (University of Greifswald and TU Darmstadt) and on an ongoing basis, he was able to build up and now contributes an extensive network to global biotechnology companies and stakeholders.

Joseph Heenan (CEO) has 15 years of experience managing and technically leading machine learning teams in a Fortune 100 company and has worked in three software start-ups with successful exits. Joe has extensive experience scaling software teams, rapid prototyping, and commercializing machine learning research and development.



BUSINESS IDEA

Protein Search Engine Built by and for Biologists

Identifying new proteins can lead to breakthroughs in understanding diseases, developing new treatments, and advancing biotechnology and sustainability. This knowledge enhances drug design, personalized medicine, and the development of bioengineered products, driving innovation in health and industry.

Accelerating Protein Engineering

Proteineer's mission is to accelerate biological research via organizing and annotating the entirety of the world's publicly available genomics data. The company helps teams of biologists and bioinformaticians to discover, annotate, analyze and unify genomics data from a vast amount of different sources with unprecedented thoroughness and ease. Proteineer's APIs allow for deep integration with existing systems and processes.

Trusted by Users and Partners

GeneStore is the only offering on the market that uses spot instances and cloud compute to cost-efficiently scan and annotate the petabytes of publicly-available genome data. It provides datasets and insights to customers

to help accelerate their own protein engineering and machine learning initiatives.

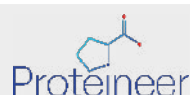
Through Proteineer's participation in various accelerators and research consortia it has already achieved widespread visibility.

In the competitive and rapidly evolving landscape of bioinformatics, Proteineer is in a unique niche by revolutionizing the protein discovery process.

Built by and for Biologists

Proteineer's business model is anchored in the Software as a Service (SaaS) paradigm, providing clients with subscription-based access to its GeneStore platform. This ensures that users continuously benefit from the latest advancements and updates without the need for significant upfront investment. The subscription model also fosters an ongoing relationship with clients, allowing Proteineer to offer consistent support and gather feedback for continual improvement.

- Proteineer, Dietzenbach, Germany
- www.proteineer.com



ELEVATOR PITCH

Driving Biotechnological Innovations

Over the past decade, life sciences have been revolutionized by affordable DNA sequencing technologies. Since 2008, the cost of sequencing DNA has dropped dramatically, making genomic data more accessible. However, the challenge has shifted to managing and interpreting the vast amounts of raw data generated. Stored in the Sequence Read Archive (SRA), this data is a treasure trove of potential discoveries but is underutilized due to its complexity. Researchers face significant challenges in processing and analyzing this data, limiting the scope of their scientific inquiries.

Proteineer, founded in January 2022, aims to empower life science researchers to fully harness DNA sequencing data. Our Software-as-a-Service (SaaS) platform allows clients to efficiently search and analyze protein data. Resulting for most queries in more than 200,000 results. We enrich this data with modern machine learning algorithms to predict properties such as thermostability or structures.

Our primary market is the industrial biotechnology sector, responsible for producing basic and specialty chemicals, pharmaceuticals, and agricultural products, with an annual revenue exceeding €300 billion. Proteineer offers unparalleled data analysis capabilities, accelerating research and development in this

robust and rapidly growing industry. Our proprietary search system allows searching for more data, giving our clients a competitive edge. This technological advantage, combined with our first-to-market status, positions Proteineer as a leader in driving biotechnological innovations and sustainable solutions across the industry.

Milestones

- 2022**
 - Q1 Founding
- 2022-**
 - Q2 First Publicly Traded Customer
- 2023**
 - Q1: Web App to display search results
 - Q4: Accepted into Hessian.AI AI Startup Rising Accelerator
- 2024**
 - Q1: Start of publicly funded CO₂BioTech Project for CO₂-based bioproduction platform for cysteine, aspartate, and glycolate.
 - Q3: Docking Based search

Roadmap

- 2023-Q4 In silico protein generation based on search results.

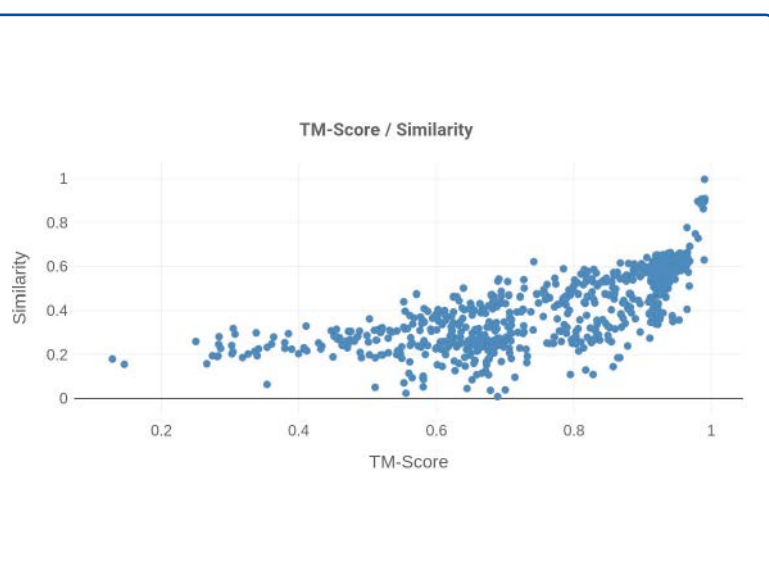


Fig. 1: The GeneStore not only searches for sequence similarity but as well for structural similarity of proteins. The plot shows how sequences with a high structure similarity, yet a low similarity on the sequence are found.

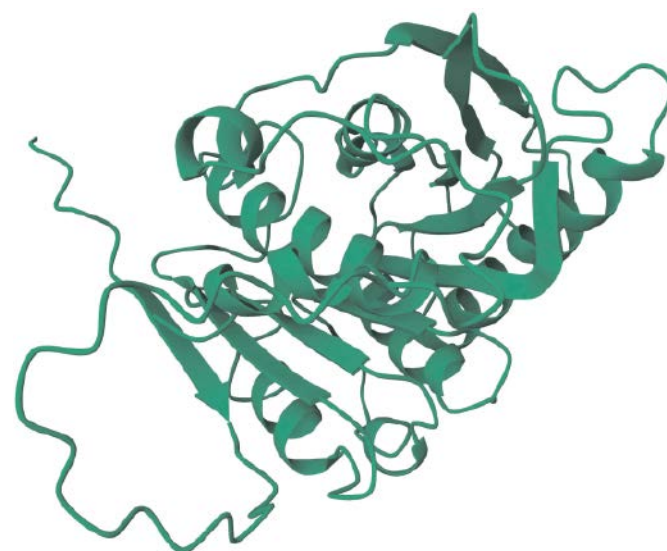


Fig. 2: The GeneStore predicts structures for all hits, many of which cannot be found elsewhere, enabling even better cherry picking.